Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes

**ELISABETH GRUNDNER-CULEMANN, GLOVER W. MARTIN III, JOHN W. HARNEY,
and MARLA J. BERRY**

Thyroid Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School,
Boston, Massachusetts 02115, USA

ABSTRACT

Translation of UGA as selenocysteine requires specific RNA secondary structures in the mRNAs of selenoproteins. These elements differ in sequence, structure, and location in the mRNA, that is, coding versus 3′ untranslated region, in prokaryotes, eukaryotes, and archaea. Analyses of eukaryotic selenocysteine insertion sequence (SECIS) elements via computer folding programs, mutagenesis studies, and chemical and enzymatic probing has led to the derivation of a predicted consensus structural model for these elements. This model consists of a stem-loop or hairpin, with conserved nucleotides in the loop and in a non-Watson–Crick motif at the base of the stem. However, the sequences of a number of SECIS elements predict that they would diverge from the consensus structure in the loop region. Using site-directed mutagenesis to introduce mutations predicted to either disrupt or restore structure, or to manipulate loop size or stem length, we show that eukaryotic SECIS elements fall into two distinct classes, termed forms 1 and 2. Form 2 elements have additional secondary structures not present in form 1 elements. By either insertion or deletion of the sequences and structures distinguishing the two classes of elements while maintaining appropriate loop size, conversion of a form 1 element to a functional form 2-like element and of a form 2 to a functional form 1-like element was achieved. These results suggest commonality of function of the two classes. The information obtained regarding the existence of two classes of SECIS elements and the tolerances for manipulations of stem length and loop size should facilitate designing RNA molecules for obtaining high-resolution structural information about these elements.

Keywords: 3′ untranslated region; recoding; RNA structure; translation

INTRODUCTION

Interactions between specific RNA secondary structures and their cognate RNA binding proteins mediate a wide variety of catalytic and regulatory processes, from RNA splicing and editing to development-specific RNA localization and translational control of gene expression. Translation of UGA codons as selenocysteine, a selenium analog of cysteine, requires specific secondary structures present in the mRNAs encoding selenoproteins. This recoding process, whereby a stop codon is redirected to specify incorporation of this unusual amino acid, occurs in a few select proteins in prokaryotes, archaea, and eukaryotes. In all three lines of descent, the requisite structures are stem-loops, but the identities and positions of conserved nucleotides

and the predicted secondary structures are quite distinct. In addition, the positions of the structures within the mRNAs differ, being found in the coding region in prokaryotes, and in the 3′ untranslated regions (UTRs) in eukaryotes and archaea.

In prokaryotes, the secondary structures that direct selenocysteine incorporation have been shown to serve as recognition sites for binding of a selenocysteyl-tRNA-specific translation elongation factor. This elongation factor binds the mRNA stem-loops and the charged tRNA, and is envisioned to deliver the tRNA to the UGA selenocysteine codon at the ribosome. Despite considerable efforts, much less is known about the mechanism of selenocysteine incorporation in eukaryotes. We are investigating this process, and the stem-loop structures, termed SElenoCysteine Insertion Sequence, or SECIS elements, that direct it. We are interested in the specific structural features of these elements required for function, presumably via their interaction with a putative eukaryotic selenocysteyl-

tRNA-specific elongation factor. No such factor has been functionally identified to date, but much effort has gone into identifying candidate proteins using SECIS element binding as a criteria (Shen et al., 1995b, 1998; Hubert et al., 1996; Lesoon et al., 1997).

Eukaryotic SECIS elements contain several highly conserved features, including the sequences AUGA and GA at the 5′ and 3′ base of the stem, respectively, which form a non-Watson–Crick G-A, A-G tandem pair (see Fig. 1). We refer to this region at the base of the stem, the most conserved region in these elements, as the SECIS core (Martin et al., 1998). A stretch of two or three adenosines is also conserved in a loop or bulge at the top of the stem. Finally, the stem separating the SECIS core from the conserved adenosines is fixed at 9–11 bp, approximately one helical turn of an A-form RNA helix. All eukaryotic SECIS elements identified to date possess these features, and mutagenesis studies have confirmed the importance of the conserved nucleotides and of base pairing in the stems for function (Berry et al., 1993; Shen et al., 1995a; Martin et al., 1996, 1998). Site-directed mutagenesis studies by our group and others have focused on the SECIS elements from a few specific selenoproteins, the rat type 1 deiodinase (D1; Berry et al., 1993; Martin et al., 1996, 1998), the rat and human cytoplasmic glutathione peroxidase (cGPX; Berry et al., 1993; Shen et al., 1995a; Martin et al., 1998; Walczak et al., 1998), and rat selenoprotein P (selP; Berry et al., 1993). Chemical and enzymatic probing to investigate the secondary structures of SECIS elements also utilized only the D1 and

cGPX elements (Walczak et al., 1996, 1998). These studies, combined with computer modeling, resulted in derivation of a predicted secondary and tertiary structure for the consensus SECIS element (Walczak et al., 1996). However, closer scrutiny reveals striking differences between SECIS sequences that allow the elements to be divided into two distinct classes of structures. These classes are distinguished by the presence or absence of a predicted additional small stem-loop at the top of the SECIS element, above the conserved adenosines. In these structures, the conserved adenosines are predicted to reside in an internal bulge, rather than a terminal loop. We previously termed these classes form 1 and form 2, the latter bearing the predicted additional secondary structure (Low & Berry, 1996). The form 1 class includes the D1, cGPX, gastrointestinal GPX, and the second selP SECIS elements. The form 2 class includes all others, the first selP element, and those of the type 2 and 3 deiodinases, plasma and phospholipid hydroperoxide GPXs, selenoprotein W, selenophosphate synthetase, thioredoxin reductase, and the recently identified 15-kDa prostate selenoprotein (Table 1).

As the form 1 versus form 2 distinction is based on computer predictions of folding patterns, we sought to determine whether these predicted structures actually form in vivo, and if so, whether the additional base pairing is required for function. We investigated the effects of converting a form 2 element to form 1, and a form 1 element to form 2. Because of the significant size differences in the SECIS loops in the two classes
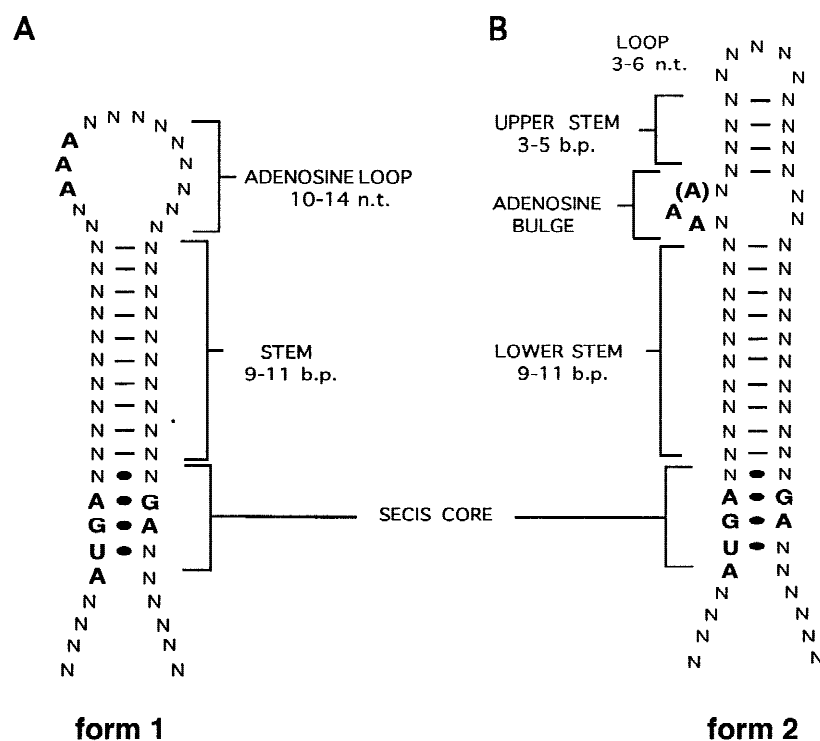

FIGURE 1. Diagram of consensus form 1 and form 2 classes of SECIS elements. **A**: Form 1 element. Form 1 elements consist of a 9–11-bp stem separating the conserved SECIS core from a 10–14-nt terminal loop containing three adenosines on the 5′ side of the loop. **B**: Form 2 element. Form 2 elements differ from form 1 elements in that the conserved adenosines are present in an internal bulge, separated from a 3–6-nt terminal loop by a 3–5-bp upper stem. Conserved nucleotides are shown in bold.

TABLE 1. FORM 1 AND FORM 2 EUKARYOTIC SECIS ELEMENTS

FORM 1

	Core	Stem	Loop	Stem	Core	Stem length	Loop size
Type 1 deiodinase							
rat	AUGAU	GGUCACAGU	GUAAAGUUCACACA	GCUGUGACU	UGAU	9	14
mouse	AUGAU	GGUCACAGU	GUAAAGUUCCCACA	GCUGUGACU	UGAU	9	14
human	AUGAU	GGCCACAGCC	UAAAGUACACAC	GGCUGUGACU	UGAU	10	12
canine	AUGAC	GGCCACAGC	CUAAAGCACACACA	GCUGUGACU	UGAU	9	14
Cytoplasmic glutathione peroxidase							
rat	AUGAC	GGUGUUUCCUC	UAAAUUUACAU	GGAGAAACACC	UGAU	11	11
mouse	AUGAU	GGUGUUUCCUC	UAAAUUUGCAC	GGAGAAACACC	UGAU	11	11
human	AUGAG	GGUGUUUCCUC	UAAACCUACGAG	GGAGGAACACC	UGAU	11	12
bovine	AUGAA	GGUGUUUCCCUC	UAAACCUACGU	GGAGGAAUGCC	UGAU	11	11
rabbit	AUGAG	GGCGUUCCCCC	GAAAACAAAU	GGAGGAACGCC	UGAU	11	10
S. mansoni	AUGAC	GAUGGCAAUC	UCAAAUGUUCAUU	GGUUGCCAUU	UGAU	10	13
Gastrointestine specific glutathione peroxidase							
human	AUGAU	GGCACCUUCC	UAAACCCUCAU	GGGUGGUGUC	UGAG	10	11
Selenoprotein P element 2							
rat	AUGAC	GGUUUAAUAG	AGAAACUGAGUC	CUAUGAACC	UGAA	10	12
mouse	AUGAU	GGUUUCAUAG	AGAAACUAAGUC	CUAUGAACC	UGAC	10	12
human	AUGAU	GGUUUAAUAGG	UAAACCAAAC	CCUAUAAACC	UGAC	11	10
bovine	AUGAU	GGUUCAAUAGG	CAAACCAGGU	CCUAUAAACC	UGAA	11	10

FORM 2

	Core	Lower stem	Adenosine bulge	Upper Stem	Loop	Upper Stem	Bulge	Lower stem	Core	Stem length	"Form 1 Loop"	Form 2 Upper Stem	Form 2 Loop
Type 2 deiodinase													
human	AUGAU	AACUACUGAC	GAAA	GAGU	CAUCG	ACUC	A	GUUAGUGGUU	GGAU	10	18	4	5
mouse	AUGAU	AACUACUGAC	GAAA	GAGC	UGUCU	GCUC	A	GUCUGUGGUU	GGAU	10	18	4	5
Type 3 deiodinase													
human	AUGAC	GAACUAUCUC	UAAC	UGGU	CUUG	ACCA	C	GAGCUAGUUC	UGAA	10	17	4	4
rat	AUGAC	GAACCGCCUC	UAAC	UGGG	CUUG	ACCA	C	GGGUCGGCUC	UGAA	10	17	4	4
Xenopus	AUGAC	GAC*CGAUUUU	GAAA	UGGU	CUCACG	GCCA	A	AAACUCGUGUC	CGAC	10	18	4	6
Plasma glutathione peroxidase													
rat	AUGAA	GGAGGGGCCC	GAA	GCCC	UUGU	GGGC		GGGCCUCCCC	UGAG	10	15	4	4
mouse	AUGAA	GGAGGGGCCC	GAA	GCCC	UUGU	GGGC		GGGCCUCCCC	UGAG	10	15	4	4
human	AUGAG	GGAGGGGCCC	AAA	GCCC	UUGU	GGGC		GGACCUCCCC	UGAG	10	15	4	4
bovine	AUGAA	GGAGGGGCCC	GAA	GCCC	GCGU	GGGC		GGGCCUCCCU	UGAG	10	15	4	4
Phospholipid hydroperoxide glutathione peroxidase													
rat	AUGAC	GGUCUGCCU	GAAA	ACCAG	CCCG	CUGGU	G	GGGCAGUCC	CGAG	9	19	5	4
mouse	AUGAA	GGUCUGCCU	GAAA	ACCAG	CCUG	CUGGU	G	GGGCAGUCC	UGAG	9	19	5	4
pig	AUGAC	AGUCUGCCU	AAAA	ACCA	GCCC	UGGU	G	GGGCAGACU	CGAG	9	17	4	4
Molluscum contagiosum glutathione peroxidase													
human	AUGAC	GGCGUCUCUC	GAA	CACC	GACAA	GGAG	G	GAGAGCUGCC	CGAG	10	17	4	5
Thioredoxin reductase													
human	AUGAA	GUCACCAGUCU	CAA	GCC	CAUGU	GGU		AGGC*GGUGAU	GGAA	10	14	3	5
rat	AUGAA	GUCACUAGCCU	CAA	GCC	CAAGU	GGU		GGGC*AGUGAC	AGAA	10	14	3	5
Selenoprotein P element 1													
rat	AUGAG	AACAGAAAC	AUAAA	CUA	UGACC	UAG	GG	GUUUCUGUU	GGAU	9	18	3	5
mouse	AUGAG	AACAGAAAC	AUAAA	CUA	UGACC	UAG	GG	GUUUCUGUU	GGAU	9	18	3	5
human	AUGAG	AAUA*GAAAC	GUAAA	CUA	UGACC	UAG	GG	GUUUUCUGUU	GGAU	9	18	3	5
bovine	AUGAG	AACAGAAAC	GAAAA	CUA	UAACC	UAG	GG	GUUUCUGUU	GGAU	9	18	3	5
Selenoprotein W													
rat	AUGAC	AGGAAGGACUG	AAAU	GUC	UCAAA	GAC	CUG	UGGUCUUUCUU	CGAU	11	18	3	5
mouse	AUGAC	AGGAAGGACUG	AAAU	GUC	UUA	GAC	CUG	UGGUCUUUCCU	CGAU	11	16	3	3
human	AUGAU	AGGAAGGACUG	AAAA	GUC	UUGUG	GAC	ACC	UGGUCUUUCCC	UGAU	11	18	3	5
monkey	AUGAU	AGGAAGGACUG	AAAA	GUC	UUGUG	GAC	GCC	UGGUCUUUCCC	UGAU	11	18	3	5
sheep	AUGAC	AGGAAGGACUG	AAAU	GUC	UCUUG	GAC	GCC	UGGUCCUUCCU	UGAU	11	18	3	5
15 kDa prostate selenoprotein													
human	AUAAA	GACU*ACACAG	AAAA	CCUU	UCU	AGGG	AU	UUGUGUGGAUC	AGAU	10	17	4	3
mouse	AUGAG	GAUU*ACACAG	AAAA	CCUU	UGUU	AAGG	AC	UUGUGUAGAUC	UGAU	10	18	4	4
Selenophosphate synthetase 2													
mouse	AUGAU	GUCUCUCCCU	CUAAC	UCC	CAGUAA	GGA	C	UGGGAGAGGC	UGAA	10	18	3	6
human	AUGAC	GUCUCUCCCU	CUAAA	CCC	CAUUAA	GGA	C	UGGGAGAGGC	AGAG	10	18	3	6

Asterisks indicate gaps on one side of the stem opposite unpaired nucleotides on the other side. Underlines indicate potential base pairing in upper stem.

of elements, we examined the effects of manipulating the sizes of both a form 1 and a form 2 SECIS loop on activity.

Finally, we turned our attention to possible constraints on stem length. We previously showed through mutations disrupting Watson–Crick interactions, and compensatory mutations restoring these interactions, that base pairing in the SECIS stem was required for function (Berry et al., 1993; Martin et al., 1996). Because of the narrow range of stem lengths found in nature, we examined the effects of shortening or lengthening the stem on function. The implications of our findings from these studies for form 1 and 2 SECIS elements are discussed.

RESULTS

SECIS elements fall into two distinct structural classes

The sequences of all eukaryotic SECIS elements identified to date contain the conserved nucleotides 5′AUGA3′ and 5′GA3′, forming a non-Watson–Crick base paired core at the base of the SECIS stem (Table 1 and Figs. 1A and 1B, red nucleotides). A stretch of two to three adenosines is also conserved (Table 1, blue nucleotides). Assuming that the 9–11-bp stems above the core all end in open loops with no additional secondary structure, the loop sizes would range from 10–19 nt. While loops in the range of 4–12 nt are predicted to be thermodynamically stable when closed by a stable stem, loops of, for example, 16 nt or larger are considerably less stable (Turner & Sugimoto, 1988). Thus, additional base pairing interactions would likely form within the larger loops to stabilize these structures. We examined, by computer predicted or manipulated folding, the potential for additional secondary structures in those SECIS elements with large loop sizes, and found that virtually all could form similar structures. These structures consisted of three to five Watson–Crick base pairs forming an "upper stem" immediately above the conserved adenosines, closing a 3–6 nt terminal loop (Table 1). We have termed these elements with additional secondary structure near the top form 2 SECIS elements (Low & Berry, 1996). A consensus form 2 structure is shown in Figure 1B. The internal bulge of conserved adenosines on the left (5′) side of the stem (typically 3–5 nt) between the upper and lower stems is typically found opposite a smaller bulge (usually 1–3 nt) on the right (3′) side, although in some cases no unpaired nucleotides are present. The classical SECIS element containing a terminal loop of 10–14 nt is shown for comparison (Fig. 1A, form 1). In these elements, no potential for additional Watson–Crick base pairing in the loops is predicted. Even if the possibility of secondary structures with nonstandard base pairing were considered, secondary structures of the form 2 type would be very unlikely to form, because the small number of nucleotides in the form 1 loops would result in either the stems or the loops being too small to be stable. Based on these differences, we separated the known elements into two groups, those with terminal loops of 14 nt or less (Table 1, Form 1 Class), and those with loops of 15 nt or greater (Table 1, Form 2 Class).

Effects of additional secondary structure in a form 2 SECIS element on function

The additional secondary structures predicted for the form 2 SECIS elements are likely to form in vivo, but this folding might be fortuitous. Any actual requirement of these predicted structures for function has not been reported to date. We chose to investigate this require-

ment in the first SECIS element in rat selP. We utilized a previously characterized reporter system for assessing SECIS function (Martin et al., 1996). In this system, various wild-type or mutant SECIS elements are subcloned downstream of the D1 coding region in a mammalian expression vector. The ability of the SECIS elements to direct selenocysteine incorporation is assessed by measuring the level of D1 activity produced. No D1 activity is produced in the absence of a functional SECIS element, and this system allows reproducible comparison of various activity levels relative to a positive control as standard. The minimal wild-type selP first SECIS element was constructed and tested for activity, relative to the minimal D1 element. The predicted structure of the selP element is shown in Figure 2A. The activity levels of these two elements were found to be equivalent (Fig. 2C). This is in agreement with our previous findings for other SECIS elements, showing that form 1 and form 2 elements exhibit comparable activity levels (Martin et al., 1998).

To test whether the additional base pairing in a form 2 element affects SECIS function, the region consisting of the predicted upper stem base pairs in the first selP element was mutagenized to eliminate pairing. This upper stem consists of three predicted base pairs. We converted either the first 2 or all 3 nt on the 3′ half of the stem, 5′UAG3′, to the complementary sequences, 5′AUC3′ or 5′AUG3′, disrupting either the first 2 or all 3 bp. The predicted disrupted structure is shown in Figure 2A. Disrupted base pairing would generate a thermodynamically unfavorable, open 18-nt loop. Both of these changes resulted in complete loss of activity (Fig. 2C). While the 5′AUG 3′ mutation could potentially fold into an alternative form 2-like structure (base pairing between 5′CUAU 3′ and 5′AUGG 3′ forming the upper stem), other unfavorable interactions in this region are also possible, and the actual folded structure assumed in these mutant constructs is not known.

We next examined the effects of introducing compensatory substitutions in the 5′ half of the upper stem, which would restore base pairing. The 3 nt, 5′CUA3′, were converted to the complementary sequence, 5′GAU3′, in the context of the disrupting mutation, 5′AUC3′, on the 3′ half of the stem (Fig. 2A). Restoring the possibility of base pairing resulted in restoration of 31% of activity, indicating that the secondary structure in this region contributes significantly to function. However, the lack of full restoration of activity indicates either that some sequence specificity is required, or that the structure of the disrupted and restored stem is different from that of the wild-type element.

Conversion of a form 2 element to a form 1 element

To further investigate whether the additional secondary structure in a form 2 element was required for func-
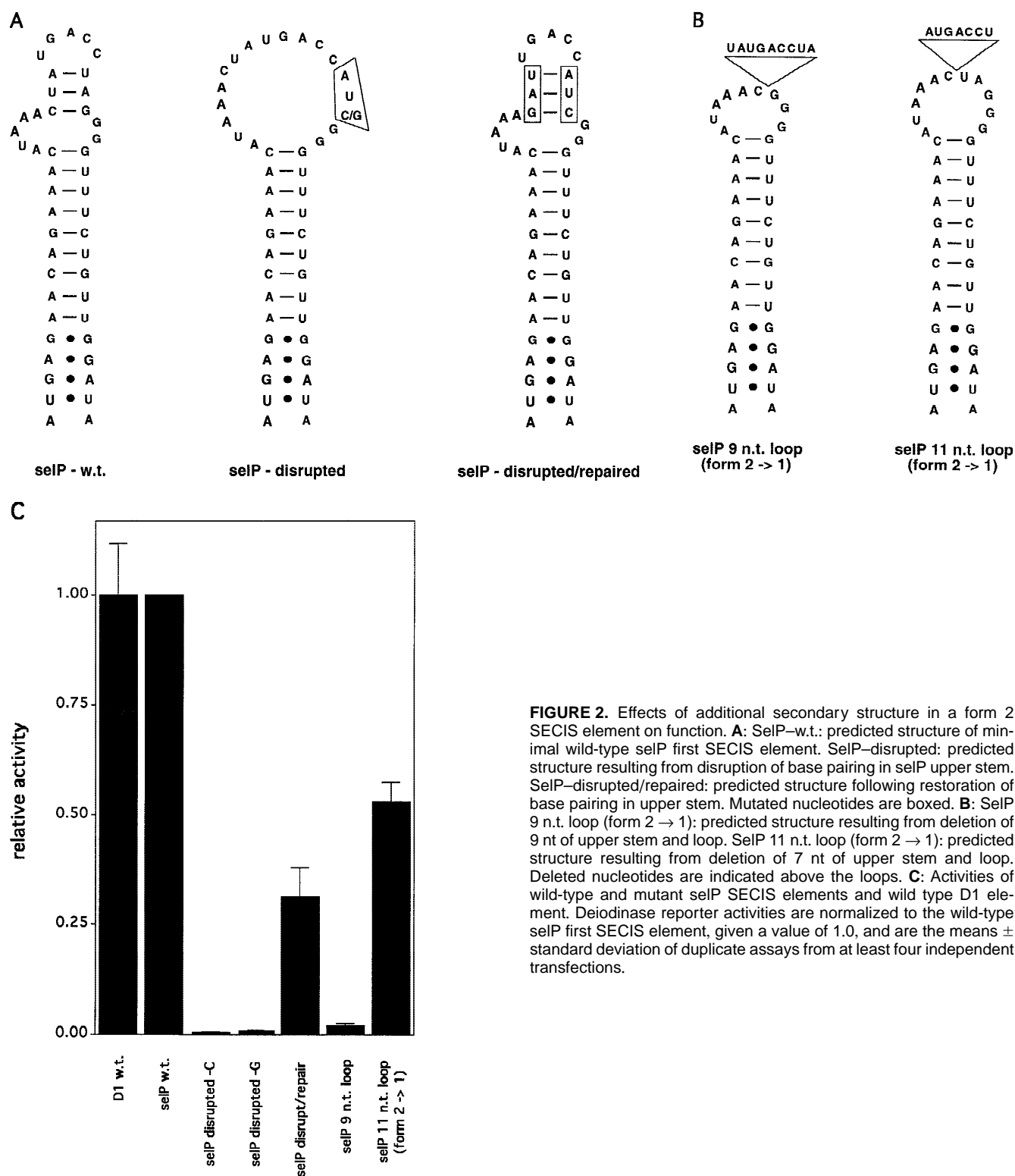
FIGURE 2. Effects of additional secondary structure in a form 2 SECIS element on function. **A**: SelP–w.t.: predicted structure of minimal wild-type selP first SECIS element. SelP–disrupted: predicted structure resulting from disruption of base pairing in selP upper stem. SelP–disrupted/repaired: predicted structure following restoration of base pairing in upper stem. Mutated nucleotides are boxed. **B**: SelP 9 n.t. loop (form 2 → 1): predicted structure resulting from deletion of 9 nt of upper stem and loop. SelP 11 n.t. loop (form 2 → 1): predicted structure resulting from deletion of 7 nt of upper stem and loop. Deleted nucleotides are indicated above the loops. **C**: Activities of wild-type and mutant selP SECIS elements and wild type D1 element. Deiodinase reporter activities are normalized to the wild-type selP first SECIS element, given a value of 1.0, and are the means ± standard deviation of duplicate assays from at least four independent transfections.

tion, we deleted the top 2 bp of the upper stem and the terminal loop of the first selP element, generating a form 1-like element (Fig. 2B). The resulting element has a predicted 9-nt loop. This deletion resulted in only 2% of the wild-type level of activity. However, during the course of these studies, we determined that a 9-nt loop is below the limits tolerated for a form 1 element (see below). Therefore, we next deleted only the top base pair

of the upper stem and the terminal loop, generating a form 1-like element with an 11-nt loop (Fig. 2B). This resulted in 53% activity (Fig. 2C). Thus, a functional form 1 element can be generated from a form 2 element, provided a minimal loop size is maintained.

Conversion of a form 1 element to a form 2 element

Extending this line of reasoning, we next investigated the ability to introduce additional secondary structure into a form 1 element, generating a form 2-like element. As shown in Figure 3A, two distinct form 2-like structures were generated from the form 1 D1 SECIS element. In the first structure, a 9-nt substitution in place of five wild-type nucleotides in the loop introduces a 3-bp upper stem and a 5-nt loop. The second structure was generated fortuitously upon introduction of 3 bp into the D1 stem (Fig. 3A). The introduced nucleotides are again predicted to form a 3-bp upper stem and a 5-nt loop, with a slightly larger adenosine bulge. In both cases, the introduced sequences resulted in activity levels in the range of the wild-type D1 element (Fig. 3B). In contrast, introduction of additional nucleotides into the loop without introducing these predicted secondary structures greatly decreased function (see below).

Effects of D1 SECIS loop changes on function

The size range of putative form 1 loops in nature is 10–14 nt (see Table 1), with the D1 SECIS loop being at the high end of this range. The form 2 loops, in the absence of predicted additional secondary structure, would fall into a range of 14–19 nt. We investigated the effects of manipulating the D1 loop size on SECIS function, either by introducing insertions or generating progressive deletions (Fig. 4). As we have previously shown the conserved adenosines to be crucial for function (Berry et al., 1993), we did not make any alterations in this region of the loop, but rather, began 2 nt 3′ of the last conserved adenosine. Increasing the loop size by 3 nt, without introducing the potential for additional secondary structure, decreased activity by 73% (Fig. 4B). This contrasts with the wild-type activity levels obtained in two different constructs upon increasing the loop size by 4 nt, while generating the potential to fold into a form 2-like element (Fig. 3A). Thus, increasing the loop size without secondary structure formation appears to be detrimental to function.

We next turned to decreasing loop size. The first 2-nt loop deletion (Fig. 4A, UU deletion) resulted in a 38% increase in activity compared to wild-type (Fig. 4B). Intriguingly, this brings the D1 loop closer to the size range of the other form 1 elements. To determine if this increase in activity was a general effect of the smaller
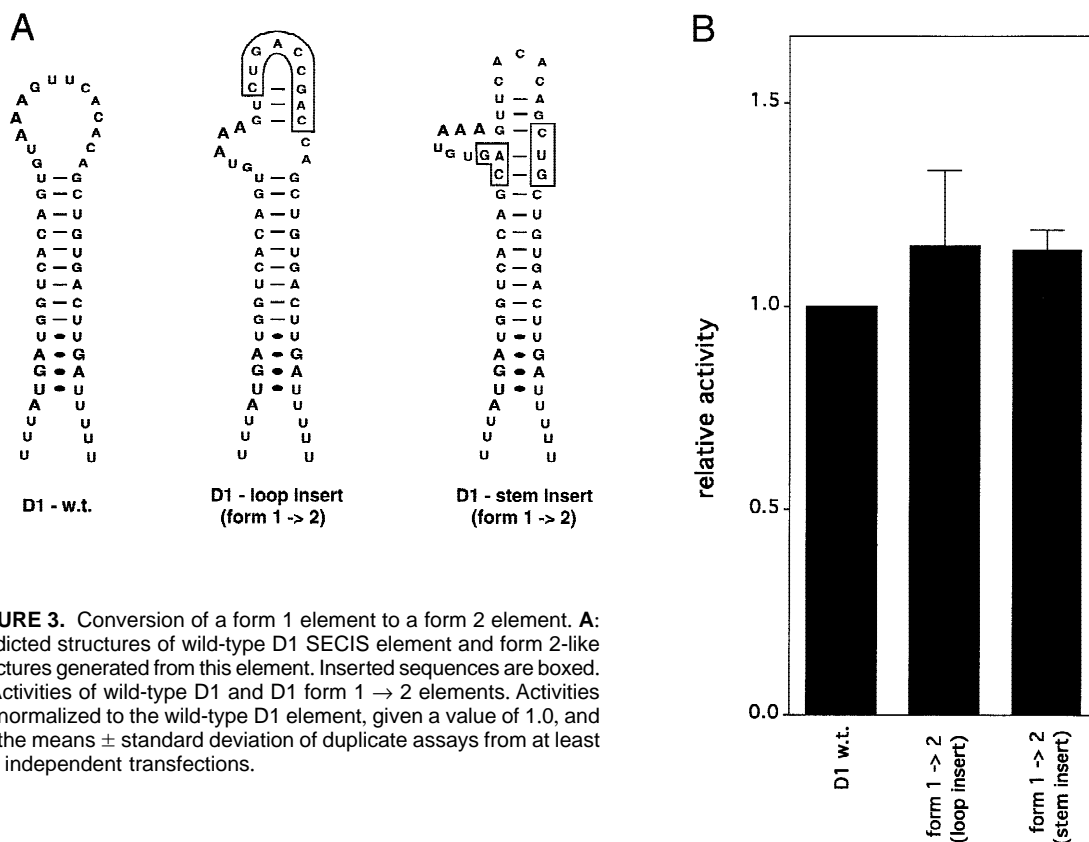

FIGURE 3. Conversion of a form 1 element to a form 2 element. **A**: Predicted structures of wild-type D1 SECIS element and form 2-like structures generated from this element. Inserted sequences are boxed. **B**: Activities of wild-type D1 and D1 form 1 → 2 elements. Activities are normalized to the wild-type D1 element, given a value of 1.0, and are the means ± standard deviation of duplicate assays from at least four independent transfections.
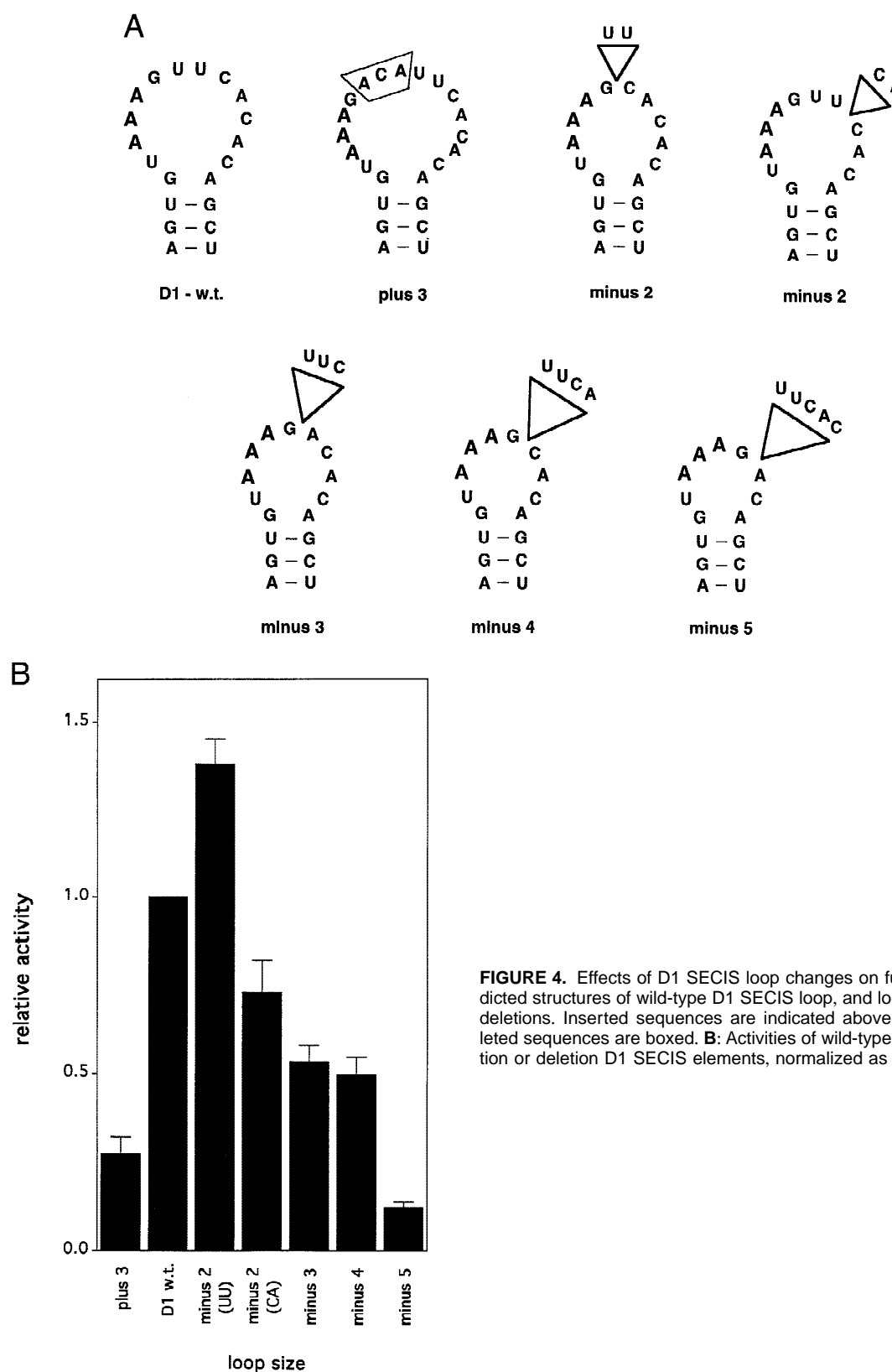
FIGURE 4. Effects of D1 SECIS loop changes on function. **A**: Predicted structures of wild-type D1 SECIS loop, and loop insertions or deletions. Inserted sequences are indicated above the loops. Deleted sequences are boxed. **B**: Activities of wild-type and loop insertion or deletion D1 SECIS elements, normalized as in Figure 3B.

size, versus a sequence-specific effect, we generated a second 2-nt deletion (CA deletion). This produced a 28% decrease in function. The disparate results with these two deletions suggests that there may be both sequence-specific and loop-size effects on function. Further loop deletions of 3 or 4 nt (generating 11- or 10-nt loops) reduced activity to the 50% range (Fig. 4B). This is similar to the activity attained with conversion of the

form 2 selP loop to an 11-nt form 1-like loop (see Fig. 2C). While both of these elements retained significant function, their sequences would preclude forming any additional secondary structures within the loops (Fig. 4A), arguing against any possible form 2-like structure forming. The 5-nt deletion (9-nt loop) resulted in an 88% loss of function. This is consistent with our findings with the first selP SECIS element, in which deletion to a 9-nt loop resulted in a 98% loss of function. Strikingly, the smallest naturally occurring form 1 loop is 10 nt.

Effects of D1 SECIS stem-length changes on function.

The predicted stem lengths for all SECIS elements fall within a very narrow range of 9–11 bp. As the length required for one helical turn of an RNA A-form helix is 11 bp, we reasoned that this narrow range might be due to constraints to maintain the conserved AUGA and AAA motifs on the same face of the helix, one helical turn apart. To investigate this, we tested the effects of manipulating the stem length on SECIS function. The D1 SECIS element stem falls on the short end of the range, with a length of 9 bp. Therefore, we inserted 1, 2, 3, 4, or 6 bp into this element, 1 bp from the top of the stem, increasing the stem length to a range

of 10–15 bp (Fig. 5). Increasing the stem length by 1 or 2 bp decreased activity by 60 and 65%, respectively. Increasing the length by 3, 4, or 6 bp abolished activity. The exception to this was the construct in which 3 bp were introduced into the stem, resulting in the ability to fold into a form 3 element. Next we deleted 1, 2, or 3 bp from the D1 SECIS stem, decreasing the length to 8, 7, or 6 bp, respectively. The first deletion resulted in a 43% decrease in activity, whereas the latter two completely abolished function. Thus, the constraints on stem length may in fact serve to maintain appropriate juxtaposition of the conserved sequence motifs.

DISCUSSION

The mechanism of selenocysteine insertion provides another fascinating example in a growing list of critical control steps in gene expression brought about by specific RNA–protein interactions. Information about the specific aspects of SECIS RNA structure required for protein interaction and SECIS function are crucial for identifying the protein or proteins involved, and elucidating this interesting translational mechanism. Since identification of the first eukaryotic SECIS elements, those of the type 1 deiodinase and cytoplasmic GPX (Berry et al., 1991), considerable efforts have gone into sequence and structural studies of these elements
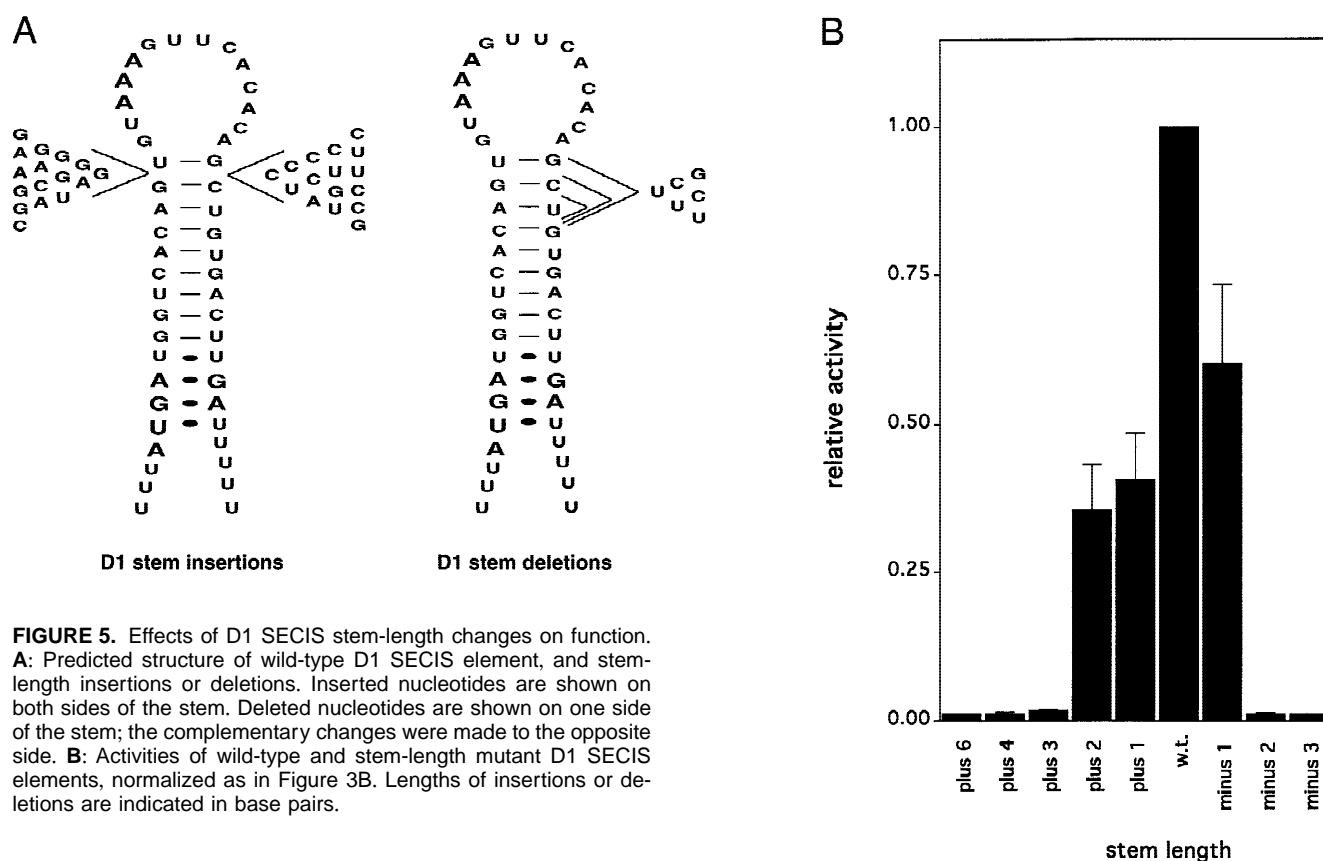

FIGURE 5. Effects of D1 SECIS stem-length changes on function. **A**: Predicted structure of wild-type D1 SECIS element, and stem-length insertions or deletions. Inserted nucleotides are shown on both sides of the stem. Deleted nucleotides are shown on one side of the stem; the complementary changes were made to the opposite side. **B**: Activities of wild-type and stem-length mutant D1 SECIS elements, normalized as in Figure 3B. Lengths of insertions or deletions are indicated in base pairs.

(Berry et al., 1993; Shen et al., 1995a; Kollmus et al., 1996; Martin et al., 1996, 1998; Walczak et al., 1996, 1998). The first SECIS elements identified were members of the form 1 class. Perhaps because they were identified first, virtually all subsequent studies focused on these elements. It is now apparent, however, that the form 1 elements constitute the minor class, with only four members identified, versus 10 in the form 2 class. Thus, the previously predicted secondary and tertiary structure models based on form 1 elements (Walczak et al., 1996, 1998; Martin et al., 1998) are only consistent with these few sequences, and a new model must be derived to accommodate the upper structure in the form 2 elements. Our predicted model for the form 2 structure is shown in Figure 1B. It includes all of the conserved features shared with the form 1 elements, and incorporates the differences unique to the form 2 elements.

The sequences of the form 1 elements apparently preclude the ability to form stable upper structures similar to those predicted for the form 2 elements. Consistent with this, Walczak et al. (1996) found that the D1 and GPX SECIS loop sequences were accessible to single-strand specific chemical and enzymatic reagents, arguing against secondary structure in this region of these elements. However, the possibility of less stable or perhaps previously unrecognized structural motifs could not be excluded, based on the available information. The results with the D1 loop deletions further argue against this, as this loop can be decreased to 10 nt while still retaining 50% activity. This size would be insufficient to accommodate a 3–4-nt adenosine bulge, plus 3–4 bp and a terminal loop.

Why would two classes of SECIS structures have evolved? While it is theoretically possible that the two classes of elements are recognized by two distinct proteins, the fact that the different structures nonetheless maintain the conserved elements and their positioning relative to each other suggests that invoking a second protein would not be necessary. Further, recent studies show that elements from either class can compete with each other for a limiting SECIS-specific factor in transfected cells (S.C. Low, J.W. Harney, & M.J. Berry, in prep.). Perhaps the additional secondary structure in the form 2 elements serves only to maintain thermodynamic stability or to nucleate SECIS element folding. Alternatively, the additional structure may serve to position the conserved adenosines appropriately such that they are present in a small bulge on the left side of the helix, rather than a large unstructured loop. In fact, the juxtaposition of the adenosines relative to the stem, and thus also to the non-Watson–Crick core, would be predicted to be very similar in the form 1 and form 2 elements. These constraints on position may be necessary for interaction of a specific binding protein with both the non-Watson–Crick core and the adenosine bulge at the same time. While the reasons for two distinct structures of SECIS elements are not clear at this time, the conservation in sequence and relative juxtaposition of the invariant regions suggests a common mechanism or mode of function or recognition. In addition, the ability to interconvert form 1 and 2 structures further argues for a common mechanism.

All SECIS elements identified to date posses a 9–11-bp stem between the non-Watson–Crick core and the conserved adenosine bulge or loop. The 9-bp D1 stem lies on the short end of the range for stem length in naturally occurring elements. Increases of 1 or 2 bp would still lie within this range, and in fact these changes are tolerated, but with reduced activity. Longer stems are poorly functional or inactive, consistent with their absence in nature. Strikingly, a decrease in length of 1 bp is better tolerated than the increases, although no elements of this reduced length have been identified to date. However, stems shorter than this are inactive. Thus, tolerance for a range of 8–11 bp suggests that there is not a stringent requirement for maintaining the two groups of conserved nucleotides on precisely the same face of the RNA helix, but rather a more flexible range of $\pm\frac{1}{4}$ helical turn apart. This is underscored by the decrease and eventual loss in function with significantly shorter or longer stems. However, these effects should be viewed within the context of the specific element under study, as changes in the juxtaposition of the adenosine loop relative to the core may depend on the size and flexibility of the adenosine bulge and the slight variations in positions of the adenosines within different elements. With the information obtained regarding tolerance for manipulations of SECIS stem-length and loop size, NMR studies and crystallization attempts can now be focused on smaller molecules, with the goal of obtaining high resolution information about these structures.

MATERIALS AND METHODS

Constructs

Oligonucleotide primers complementary to cDNA positions 1533–1555 and 1585–1558 of rat D1 and to positions 1470–1492 and 1523–1504 of rat selP were designed with terminal *Hin*dIII and *Not*I sites, respectively. Primers were used to PCR amplify the minimal SECIS elements from the corresponding wild-type cDNAs. PCR products were digested with the appropriate enzymes and subcloned into the corresponding sites of G16-D10ΔH3 (Martin et al., 1996). Mutagenesis was carried out by amplification from wild-type templates with mutagenic internal oligonucleotides containing the desired nucleotide changes in combination with the above wild-type oligonucleotides encoding the terminal restriction sites (see Table 2). PCR products were subcloned between the *Hin*dIII and *Not*I sites of G16-D10ΔH3 as described above. Once subcloned, the amplified regions were sequenced in their entirety.

TABLE 2. PCR OLIGONUCLEOTIDE SEQUENCES

<u>SelP constructs:</u>
SelP 21 (w.t. 5'end + Hind III site) – 5' AAGCTTTAGATTGATGAGAACAGAAACATAAA 3'
SelP 22 (w.t. 3'end + Not I site) – 5' GCGGCCGCGAGCTATCCAACAGAAACCC 3'

SelP 29 (disrupted upper stem) – 5' AAGCTTTAGATTGATGAGAACAGAAACATAAACTATGACC<u>ATG</u>/CGGGTTTCTGT 3'
SelP 26 (disrupted/repaired upper stem) – 5' AGAAACATAAAC<u>A</u>TTGACC<u>A</u>TGGG 3'
SelP 28 (11 nt upper loop) – 5' AACAGAAACATAAAC (tatgaccta) GGGGTTTCTGTT 3'
SelP 30 (9 nt upper loop) – 5' AACAGAAACATAAACT (atgacct) AGGGTTTCTGTT 3'

<u>Deiodinase constructs:</u>
MB 96 (w.t. 5'end + Hind III site) – 5' CCAAGCTTTAGTTATATTTGTTTATGATGGTCAC 3'
TM107 (w.t. 5'end + Hind III site) – 5' CCAAGCTTTAGTTTATGATGGTCACAGTGTAAAG 3'
TM112 (w.t. 3'end + Not I site) – 5' CCGCGGCCGCTTTTTAAAAATCAAGTCACAGCTGTGTG 3'
TM116 (w.t. 3'end + Not I site) – 5' CCGCGGCCGCCCGACATTTTTAAAAATCAAGTCAC 3'

TM115 (form 1→2: stem insert) – 5' TTTTTAAAAATCAAGTCACAG (CAG) CTGTGTGAACTTTACA (CTG)CTGTGACCATCATAAACTA 3'
EGC7 (form 1→2: loop insert) – 5' GATGGTCACAGTGTAAAGT (CTGACCGAC - replaces tcaca) CAGCTGTGACTTGATTTTTAAAAATG 3'

EGC6 (3 nt loop insert) – 5' TCAAGTCACAGCTGTGTG AA (TGT) CTTTACACTGTGACCATC 3'
TM144 (2 nt loop del) – 5' TTTTTAAAAATCAAGTCACAGCTGTGTG (aa) CTTTACACTGTGACCATCATAAACTA 3'
EGC5 (2 nt loop del) – 5' CGACATTTTTAAAAATCAAGTCACAGCTGTG (tg) AACTTTACACTGTGACCATC 3'
EGC2 (3 nt loop del) – 5' CGACATTTTTAAAAATCAAGTCACAGCTGTGT (gaa) CTTTACACTGTGACCATCATAA 3'
EGC3 (4 nt loop del) – 5' CGACATTTTTAAAAATCAAGTCACAGCTGTG (tgaa) CTTTACACTGTGACCATCATAA 3'
EGC4 (5 nt loop del) – 5' CGACATTTTTAAAAATCAAGTCACAGCTGT (gtgaa) CTTTACACTGTGACCATCATAA 3'

TM122 (3 bp stem del) – 5' TTTTTAAAAATCAAGTCAC (agc) TGTGTGAACTTTAC (act) GTGACCATCATAAACTA 3'
MB100 (2 bp stem del) – 5' TATTTGTTTATGATGGTCAC (ag) TGTAAAGTTCACACAG (ct)GTGACTTGATTTTTAAAAATG 3'
MB101 (1 bp stem del) – 5' TATTTGTTTATGATGGTCAC (a)GTGTAAAGTTCACACAGC(t)GTGACTTGATTTTTAAAAATG 3'
EGC8 (1 bp stem insert) – 5' TATTTGTTTATGATGGTCACAG(G)TGTAAAGTTCACACAG(C)CT GTGACTTGATTTTTAAAAATG 3'
EGC9 (2bp stem insert) – 5' TATTTGTTTATGATGGTCACAG(AG)TGTAAAGTTCACACAG(CT)CTGTGACTTGATTTTTAAAAATG 3'
EGC11 (3bp stem insert) – 5' TATTTGTTTATGATGGTCACAG(TGG) TGTAAAGTTCACACAG (CCA) CTGTGACTTGATTTTTAAAAATG 3'
EGC10 (4bp stem insert) – 5' TATTTGTTTATGATGGTCACAG(ACAG)TGTAAAGTTCACACAG(CTGT)CTGTGACTTGATTTTTAAAAATG 3'
TM123 (6 bp stem insert) – 5' TTTTTAAAAATCAAGTCACAG (CGGAAG) CTGTGTGAACTTTACA (CTTCCG) CTGTGACCATCATAAACTA 3'

Nucleotide substitutions are underlined. Deleted nucleotides are shown in small letters and inserted nucleotides in capital letters, in parentheses.

Transient transfections

Transient transfections were carried out in human embryonic kidney (HEK 293) cells using the calcium phosphate method of transfection as described previously (Berry et al., 1991). Three days prior to transfection, HEK 293 cells were plated onto 60 mm culture dishes in Dulbecco's modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum. Cells were transfected with 10 μg of the pUHD10-3 based expression plasmids and 4 μg of the pUHD15 plasmid (Gossen & Bujard, 1992), which encodes a protein necessary for transcriptional activation of the pUHD10-3 promoter. To monitor transfection efficiencies, cells were cotransfected with 3 μg of an expression vector containing the human growth hormone cDNA under control of the HSV thymidine kinase promoter. Media was changed 1 day following transfection. Two days after transfection, cells were harvested, washed, and resuspended in 0.1 M potassium phosphate (pH 6.9)/ 1 mM EDTA containing 0.25 M sucrose.

Northern analysis

To verify that RNA levels were not affected by the introduced mutations, Northern analysis was performed. Total RNA was prepared from two independent transfections of three loop mutants, three stem mutants, one form 1 to form 2 mutant, and wild-type D1. RNA was analyzed by Northern blotting and hybridization with a D1 probe, and a GAPDH probe to control for RNA recoveries. RNA levels varied by less than 5% in all samples but one (stem length—3 bp), which varied by 11%.

Deiodinase assays

Cells were harvested 48 h after transfection by scraping, sonicated, and cell sonicates assayed for 5' deiodinase activity as previously described (Berry et al., 1991). Briefly, cells were harvested, washed, and resuspended in 0.1 M potassium phosphate (pH 6.9)/1 mM EDTA containing 0.25 M sucrose and 10 mM dithiothreitol. Cells were then sonicated and assayed for the ability to 5' deiodinate ^{125}I reverse T3. Reactions contained 10–250 μg of protein, 1 μM ^{125}I reverse T3, and 10 mM dithiothreitol in a reaction volume of 300 μL. Reactions were incubated at 37 °C for 30 min. ^{125}I release was quantitated as described previously (Berry et al., 1991). Deiodinase activities were calculated per microliter of cell sonicate and normalized to the amount of growth hormone secreted into the media. All constructs were tested in at least three separate transfections and deiodinase assays were performed in duplicate from each transfection.

REFERENCES

Berry MJ, Banu L, Chen Y, Mandel SJ, Kieffer JD, Harney JW, Larsen PR. 1991. Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3′ untranslated region. *Nature 353*:273–276.

Berry MJ, Banu L, Harney JW, Larsen PR. 1993. Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *EMBO J 12*:3315–3322.

Berry MJ, Kieffer JD, Harney JW, Larsen PR. 1991. Selenocysteine confers the biochemical properties of the type I iodothyronine deiodinase. *J Biol Chem 266*:14155–14158.

Gossen M, Bujard H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci USA 89*:5547–5551.

Hubert N, Walczak R, Carbon P, Krol A. 1996. A protein binds the selenocysteine insertion element in the 3′UTR of mammalian selenoprotein mRNAs. *Nucleic Acids Res 24*:464–469.

Kollmus H, Flohe L, McCarthy JEG. 1996. Analysis of eukaryotic mRNA structures directing cotranslational incorporation of selenocysteine. *Nucleic Acids Res 24*:1195–1201.

Lesoon A, Mehta A, Singh R, Chisolm GM, Driscoll DM. 1997. An RNA-binding protein recognizes a mammalian selenocysteine insertion sequence element required for cotranslational incorporation of selenocysteine. *Mol Cell Biol 17*:1977–1985.

Low SC, Berry MJ. 1996. Knowing when not to stop: Selenocysteine incorporation in eukaryotes. *Trends Biochem Sci 21*:203–208.

Martin GW, Harney JW, Berry MJ. 1996. Selenocysteine incorporation in eukaryotes: Insights into mechanism and efficiency from sequence, structure, and spacing proximity studies of the type 1 deiodinase SECIS element. *RNA 2*:171–182.

Martin GW, Harney JW, Berry MJ. 1998. Functionality of mutations at conserved nucleotides in eukaryotic SECIS elements is determined by the identity of a single non-conserved nucleotide. *RNA 4*:65–73.

Shen Q, Leonard JL, Newburger PE. 1995a. Structure and function of the selenium translation element in the 3′-untranslated region of human cellular glutathione peroxidase mRNA. *RNA 1*:519–525.

Shen Q, McQuilkin PA, Newburger PE. 1995b. RNA-binding proteins that specifically recognize the selenocysteine insertion sequence of human cellular glutathione peroxidase mRNA. *J Biol Chem 270*:30448–30452.

Shen Q, Wu R, Leonard JL, Newburger PE. 1998. Identification and molecular cloning of a human selenocysteine-insertion sequence-binding protein. *J Biol Chem 273*:5443–5446.

Turner DH, Sugimoto N. 1988. RNA structure prediction. *Ann Rev Biophys Biophys Chem 17*:167–192.

Walczak R, Carbon P, Krol A. 1998. An essential non-Watson–Crick base pair motif in 3′UTR to mediate selenoprotein translation. *RNA 4*:74–84.

Walczak R, Westhof E, Carbon P, Krol A. 1996. A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA 2*:367–379.